DOCUMENT RESUME

ED 408 299 TM 026 500

AUTHOR	Fan, Xitao; And Others
TITLE	Effects of Data Nonnormality and Other Factors on Fit Indices and Parameter Estimates for True and Misspecified SEM Models.
PUB DATE	25 Mar 97
NOTE	66p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, March 1997).
PUB TYPE	Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	*Estimation (Mathematics); *Goodness of Fit; Mathematical Models; Monte Carlo Methods; *Sample Size; Statistical Bias; *Structural Equation Models
IDENTIFIERS	*Nonnormal Distributions; *Specification Error

ABSTRACT

A Monte Carlo study was conducted to assess the effects of some potential confounding factors on structural equation modeling (SEM) fit indices and parameter estimates for both true and misspecified models. The factors investigated were data nonnormality, SEM estimation method, and sample size. Based on the fully crossed and balanced 3x3x4x2 experimental design with 200 replications in each cell division, a total of 14,400 samples were generated and fitted to SEM models with different degrees of model misspecification. The major findings are: (1) mild to moderate data nonnormality has little effect on SEM fit indices and parameter estimates; (2) estimation method has considerable influence on some SEM fit indices when the model was misspecified, primarily on those comparative model fit indices; and (3) some fit indices are susceptible to the influence of sample size, and show moderate downward bias under smaller sample size conditions. Previous studies in this area have simulated a correctly-specified true model, and fit indices were found to behave consistently under different estimation methods. That finding may need to be assessed again, because considerable discrepancy of some fit indices between the two estimation methods was observed for misspecified models. It is critical that simulation studies be conducted in the presence of model misspecification. (Contains 1 figure, 8 tables, and 54 references.) (Author/SLD)

EFFECTS OF DATA NONNORMALITY AND OTHER FACTORS ON FIT INDICES AND PARAMETER ESTIMATES FOR TRUE AND MISSPECIFIED SEM MODELS

Xitao Fan
Utah State University

Lin Wang
American College Testing

Bruce Thompson
Texas A&M University
and
Baylor College of Medicine

Running Head: SEM Estimates

Note: Please address correspondence concerning this paper to: Xitao Fan, Ph.D., Department of Psychology, Utah State University, Logan, Utah 84322-2810.

Paper presented at the annual meeting of the American Educational Research Association (*session #12.10*), Chicago, March 25, 1997.

DATE: 03/27/97

ERIC/TME Clearinghouse Report

Document Cover Sheet

TM #: 026500

Title: EFFECTS OF DATA NONNORMALITY & OTHER FACTORS ON FI

Number of Pages: 57

Publication Date: 03/25/97

Document Level: 1

Notes:

ABSTRACT

The present Monte Carlo study was conducted to assess the effects of some potential confounding factors on SEM fit indices and parameter estimates for (a) both true and misspecified models. The factors investigated were (b) data nonnormality, (c) SEM estimation method, and (d) sample size. Based on the fully crossed and balanced 3×3×4×2 experimental design with 200 replications within each cell condition, a total of 14,400 samples were generated and fitted to SEM models with different degrees of model misspecification. The major findings of the study were: (a) mild to moderate data nonnormality has little effect on SEM fit indices and parameter estimates; (b) estimation method has considerable influence on some SEM fit indices when the model was misspecified, primarily on those comparative model fit indices; and (c) some fit indices are susceptible to the influence of sample size, and showed moderate downward bias under smaller sample size conditions. Previous studies in this area have overwhelmingly simulated a correctly-specified true model, and fit indices were found to behave consistently under different estimation methods. That finding may need to be revisited because considerable discrepancy of some fit indices between the two estimation methods was observed for misspecified models, even when the degree of misspecification was quite slight. Since SEM researchers rarely are certain whether they have correctly specified their models, it is critical that simulation studies are conducted in the presence of model misspecification.

Structural equation modeling (SEM) has increasingly been seen as a useful quantitative technique for specifying, estimating, and testing hypothesized models describing relationships among a set of substantively meaningful variables. Much of SEM's attractiveness is due to the method's applicability in a wide variety of research situations, a versatility that has been amply demonstrated (e.g., Baldwin, 1989; Bollen & Long, 1993; Byrne, 1994; Jöreskog & Sörbom, 1989; Loehlin, 1992; Pedhazur & Schmelkin, 1991; SAS Institute, 1990).

Furthermore, many widely used statistical techniques may also be considered as special cases of SEM, including regression analysis, canonical correlation analysis, confirmatory factor analysis, and path analysis (Bagozzi, Fornell & Larcker, 1981; Bentler, 1992; Fan, 1996; Jöreskog & Sörbom, 1989). Because of such generality, SEM has been heralded as a unified model which joins methods from econometrics, psychometrics, sociometrics, and multivariate statistics (Bentler, 1994a). In short, for researchers in the social and behavioral sciences, SEM has become an important tool for testing theories with both experimental and non-experimental data (Bentler & Dudgeon, 1996).

Despite SEM's popularity in social and behavioral research, some thorny issues still haunt SEM applications, such as the robustness of model fit assessment and parameter estimation techniques under nonnormal data conditions, the role sample size plays in SEM model fit assessment, and the effect of different estimation methods on SEM results. In SEM application in substantive research, there are two general purposes: the

assessment of model fit, and the estimation of model parameters.

Assessment of model fit requires the researcher to evaluate the adequacy of the model in relation to the empirical data drawn from a sample. If the model is judged to be adequate, then the model will then be used to explain the substantive issues of interest. At this point, model parameters estimates often become the major focus of the research. While an SEM model with adequate fit informs the researcher about the general pattern of relationships among the variables, model parameter estimates inform about the <u>direction</u> and <u>strength</u> of such relationships among the variables.

Assessment of Model Fit in SEM

χ^2 Test as a Dichotomous Decision Process

Because SEM is used to test the fit between a theoretical model and empirical data, there must be mechanisms to inform users about the adequacy of model fit. Initially, the assessment of model fit was conceptualized as a dichotomous decision process of either retaining the null hypothesis that the model fits the data, or rejecting it. The empirical basis for such a dichotomous decision traditionally was a χ^2 test assessing the degree of discrepancy between two covariance matrices: the original sample covariance matrix and the reconstructed covariance matrix based on the specified model; a small discrepancy between the two indicates reasonable fit, while a large discrepancy indicates misfit. Although this concept of model testing in SEM may be conceptually straightforward, in practice considerable uncertainty regarding model fit often arises.

As is the case with statistical significance testing in general (Thompson, 1996), the statistical significance testing approach to model fit assessment is confounded with sample size: the power of the test increases with an increase of sample size in the analysis (i.e., χ^2 tends to increase as sample size increases). As a result, model fit assessment using this narrow approach becomes stringent when sample size is large, and lenient when sample size is small.

The null hypothesis in SEM is that the model fits the data, so contrary to most hypothesis testing situations, typically the researcher wants to see that the null hypothesis is <u>not</u> rejected in SEM applications, since the specified model represents the theoretical expectations about the data structure. However, under multivariate normality assumption, SEM usually requires a relatively large sample size in order for the results of the χ^2 test to be valid (Bentler, 1992; Boomsma, 1987; Jöreskog & Sörbom, 1989). Thus, researchers using SEM methodology are in a dilemma. On the one hand, we do not want to see the null hypothesis rejected. On the other hand, SEM requires a large sample size and that large sample size inflates the power of the χ^2 test, making it easy to reject the null hypothesis. When sample size is sufficiently large, it is not surprising to see that the χ^2 test may declare a model as having poor fit with the data, even if the reconstructed covariance matrix differs trivially from the sample covariance matrix, and the model makes strong substantive sense.

<u>Descriptive Indices for Assessing Model Fit</u>

Because of the problems related to the χ^2 test for model fit

assessment in SEM (Thompson & Daniel, 1996), a variety of indices for assessing model fit have been developed for assessing the fit between a theoretical model and empirical data. Unlike the χ^2 test, which can often be used for the inferential purpose of rejecting or retaining a model, these alternative fit indices are descriptive in nature in the sense that, typically, no inferential decision is made based on these indices--these methods are used to describe the fit, rather than to test fit statistically. The relative performance characteristics of these different fit indices and their comparability under different data conditions, however, are not yet well understood. For many practitioners who use SEM in their research, it is fair to say that there exists some confusion as to which indices to use under what various data conditions.

The main reason for this situation is that different types of fit indices were developed with different theoretical rationales, and there does not seem to exist one fit index which meets all our expectations for an <u>ideal</u> fit index (assuming there even exists a consensus of expectations for such an <u>ideal</u> fit index). Although different opinions have been expressed as to what characteristics an ideal fit index should possess (Cudeck & Henly, 1991; Tanaka, 1993), it is generally accepted that an <u>ideal</u> fit index should possess three characteristics. The index should: (a) have a range between 0 and 1, with 0 indicating complete lack of fit, and 1 indicating perfect fit; (b) be independent of sample size; and (c) have known distributional properties to assist in interpretation (Gerbing & Anderson, 1993). Although quite a few fit indices are

designed to possess the first characteristic, it is not yet fully clear which fit indices possess the second characteristic. Up to now, none of the fit indices available possess the third characteristic.

Since SEM fit indices were developed with different rationales and with different motivations (Gerbing & Anderson, 1993), they may differ on one or several dimensions. Tanaka (1993) proposed a six-dimension typology for SEM fit indices, and attempted to categorize some popular fit indices along these six dimensions. This multifaceted nature of fit indices not only makes the comparison among fit indices difficult, but also makes it very difficult to select the "best" index from all those available based on the theoretical rationales upon which they were developed.

Statistically, most popular fit indices fall into one of several types. Indices of the first type--covariance matrix reproduction indices--attempt to assess the degree to which the reproduced covariance matrix based on the specified model has accounted for the original sample covariance matrix. This type of fit index can be conceptualized as the multivariate counterpart of the coefficient of determination (R^2), as in regression or ANOVA analysis (Tanaka & Huba, 1989). Examples of this type of fit indices are the Goodness-of-Fit Index (GFI) and the Adjusted Goodness-of-Fit Index (AGFI) (Jöreskog & Sörbom, 1989).

Indices of the second type--comparative model fit indices--assess model fit by evaluating the comparative fit of a given model with that of a more restricted null model. In

practice, the null model is usually a model which assumes no relationship among the measured variables in the model, although reservations have been expressed about the appropriateness of using such null models as comparative baselines (Sobel & Bohrnstedt, 1985). Bentler and Bonnet's normed and non-normed fit indices (NFI and N_NFI), Bollen's incremental fit index (DELTA2) and one or two other indices belong to this family.

Indices of the third type--parsimony weighted indices--specifically take model parsimony into consideration by imposing penalties for specifying more elaborate models. More particularly, these fit indices consider both model fit and the degrees of freedom used for specifying the model. If good model fit is obtained at the expense of freeing more parameters, a penalty will be imposed. The reasoning underlying this type of model assessment is embedded in the long tradition of science going back to William of Occam's razor: between two models that fit data equally, the simpler model is more likely to be true, and therefore is also more likely to be replicated. Besides, statistically, better fit is always obtained when more parameters in the model are freed. The parsimony indices proposed by James, Mulaik and Brett (1982) and by Mulaik, James, Van Alstine, Bennett, Lind, and Stillwell (1989) represent this type. This type of fit indices is most useful for assessing competing theoretical models, and they are less informative in situations where only one model is being tested.

A recent development in model fit assessment makes use of the noncentrality statistic from the noncentral χ^2 distribution to

construct fit indices. Based on the sample noncentrality statistic, McDonald (1989) proposed an index of noncentrality. Bentler (1990) proposed the Comparative Fit Index (CFI) which also uses the sample noncentrality statistic. As with other fit indices proposed by Bentler, CFI assesses model fit relative to a baseline null model.

Factors Affecting SEM Analysis Results

Although early studies focused on the behavior of the χ^2 statistic under different data conditions (e.g., Boomsma, 1982), soon it became apparent that χ^2 statistic's dependency on sample size may confound the interpretation of results. Consequently, some later studies put more emphasis on descriptive model fit indices. Ideally, the extent to which a model is correctly specified or misspecified should be the primary, if not the sole, determinant for model fit assessment. In reality, there exist a few confounding factors which have potential impacts on SEM analyses. Three major confounding factors have attracted the attention of many researchers: data nonnormality, estimation methods used in SEM analysis, and sample size.

Model Specification

Because fit indices are designed to assess the fit, or lack thereof, between the theoretical model and the empirical data, it is obvious that fit indices should be sensitive to model misspecification conditions. Ideally, model misspecification should be the most important factor affecting SEM fit indices. The sensitivity of some fit indices to model misspecification has been examined in a few studies (Bentler, 1990; Fan, Wang, &

Thompson, 1996; La Du & Tanaka, 1989; Marsh, Balla, & McDonald, 1988). The study by Marsh et al. (1988) examined a variety of fit indices, but the extremely small number of replications in each cell condition (n=10) might have considerably limited the generalizability of conclusions from the study. One finding from the study was that the comparative model fit indices, such as NFI, tended to be non-comparable across different studies or different data sets, since their values not only depended on model specification, but also, or more importantly, depended on how bad was the null model itself.

Some other studies (Bentler, 1990; La Du & Tanaka, 1989) involved fewer indices, making performance comparison among fit indices difficult. The study by Fan et al. (1996) examined most available fit indices which are reasonably comparable. The results of the study indicate that (a) for misspecified models, the estimation method may considerably influence some fit indices, contrary to some conclusions based only on correctly-specified models (e.g., Wang, Fan, & Willson, 1996); (b) some fit indices appear to be more sensitive to model misspecification than others.

Fan et al. (1996) further pointed out that research is conspicuously lacking for misspecified models, because most previous studies focused on correctly-specified models only. As a result, the behaviors of SEM fit indices under misspecified model conditions, and the sensitivity of the fit indices to model misspecification conditions, are largely unknown. Yet, in practice most SEM researchers do not know for a certainty that the models they are investigating have been specified correctly.

Data Normality

Multivariate normality is an important consideration in multivariate methods in general, and SEM in particular. Maximum likelihood (ML) and generalized least squares (GLS) are widely used normal-theory estimation procedures in SEM. For these estimation methods, deviation from multivariate normality may yield misleading results. In the real world, however, SEM has often been applied to data not characterized by normal distributions (Bentler, 1994b; Bentler & Dudgeon, 1996; Micceri, 1989).

A review of relevant literature (Wang et al., 1996) indicates that the concern over the possible consequences of data nonnormality has led to research in two directions. The first research direction involves developing estimation procedures or test statistics that are less sensitive to or correct for data nonnormality, e.g., the asymptotically distribution free (ADF) estimation method (Browne, 1984), scaled test statistics (Chou, Bentler, & Satorra, 1991), elliptical estimators (Bentler 1983; Browne, 1984), and the heterogeneous kurtosis method (Kano, Berkane, & Bentler, 1990). Although the progress in this direction is encouraging, these alternative estimation procedures or new test statistics are more complicated and more difficult to use.

The second direction of research focuses on the robustness of normal theory methods to data normality violations. The research in this direction provides important insights about the potential consequences when data in analyses are not normal. Typically,

Monte Carlo studies were conducted to assess the consequences of data nonnormality (Bollen & Stine, 1992; Boomsma, 1982; Chou, Bentler, & Satorra, 1991; Ichikawa & Konishi, 1995; Mooijaart, 1985). As pointed out by Bentler (1994b), "asymptotic robustness theory promises to extend the range of applicability of the computationally simpler ML and GLS estimators to situations where the more difficult distribution-free methods might seem to be needed" (p. 240).

Overwhelmingly, the studies in this area focused on the performance of the χ^2 test statistic, and "very few studies are available to evaluate the performance of other fit indices when models are fitted to nonnormal data" (Wang et al., 1996, p. 231). The present study follows the second research direction in dealing with data nonnormality, i.e., to examine the robustness characteristics of SEM fit indices in nonnormal data conditions. In addition to the χ^2 test, the study examines the behavior of other SEM fit indices as well.

Estimation Methods

Relatively little is known about the influence of normal theory estimation methods on fit indices. In a few studies which examined the issue (La Du & Tanaka, 1989; Maiti & Mukherjee, 1991; Wang et al., 1996), maximum likelihood (ML) and generalized least squares (GLS) estimation procedures were used. Estimation procedures were shown to influence the value of the fit indices. But in these studies, typically very few fit indices were examined, and the performance of many other indices were unknown.

The study by Fan et al. (1996) covered more fit indices, and

the results indicated that, for misspecified models, estimation methods seem to have considerable influence on most fit indices. This, again, contradicts some tentative conclusions from studies which only examined correctly-specified model condition (e.g., Wang et al., 1996). The discrepant results between correctly-specified and misspecified models highlight the point that model specification should be one important variation considered in simulation studies in the future.

Sample Size

It is not clear how large a sample should be in SEM applications. The research findings on this issue are inconclusive (MacCallum, Roznowski, & Necowitz, 1992; Tanaka, 1987). It has been reported that small sample size led not only to untrustworthy fit indices and estimation results, but also to high rates of improper solutions occurring in simulations (Ichikawa & Konishi, 1995). A sample size of 200 in SEM applications has been considered as being relatively small by some (Boomsma, 1982; Camstra & Boomsma, 1992; Ichikawa & Konishi, 1995; MacCallum et al., 1992; Rhee, 1993). Some researchers even consider sample sizes in the thousands to be required (e.g., Hu, Bentler, & Kano, 1992; Marsh et al., 1988).

Realistically, however, such large sample sizes are often beyond the reach of researchers. It has also been noted that using a single value to delineate small from large samples is unreasonable, because models and the number of freed parameters vary from application to application. As a result, consideration of sample size should be related to model complexity and the

number of free parameters (MacCallum et al., 1992; Tanaka, 1987).

Invariably, simulation studies have investigated the behaviors of model fit indices under different sample size conditions (Anderson & Gerbing, 1984; Bearden, Sharma, & Teel, 1982; Bentler, 1990; Bollen, 1986, 1989; Fan et al., 1996; La Du & Tanaka, 1989; Marsh et al., 1988; Wang et al., 1996), because this has been considered a major weakness of the χ^2 test in SEM, and consequently, a major concern regarding the newer alternative model fit indices. The majority of fit indices investigated, including the normed-fit-index (NFI), the goodness-of-fit index (GFI), and the adjusted goodness-of-fit index (AGFI), were shown to be influenced by sample size to different degrees.

But since different indices were involved in different studies, a performance comparison of the indices across different simulation designs becomes difficult. Also, most studies looked at the earlier fit indices, such as GFI, AGFI, NFI, and some newer indices, such as McDonald centrality, Bollen's Delta2, have only rarely been investigated. Although previous studies have added to our understanding about the impact of data nonnormality and other factors in SEM applications, much still remains to be learned about the asymptotic robustness theory (Bentler, 1994b).

First, typically the χ^2 statistic has received the most attention. Given the sensitivity of χ^2 test to sample size, and the variety of other fit indices proposed for assessing SEM model fit, it is important to understand how these SEM fit indices will perform under nonnormal data conditions and some other factors. Few studies along this line are available.

Second, simulation studies in this area typically fitted true SEM models to nonnormal data, but have rarely used misspecified models. Under true model, many sample fit indices have a ceiling effect of about 1.00, and such ceiling effects may have masked some potential differences between estimation methods (ML versus GLS), and performance differences among different fit indices. Some research related to misspecified SEM models indicate that this concern has some empirical support (Fan et al., 1996). To increase our understanding of SEM fit indices, the present study had the following research objectives:

1. to assess the impact of data nonnormality on SEM fit indices and SEM parameter estimates;
2. to assess the sensitivity of different SEM fit indices to model misspecification conditions;
3. to assess how normal theory estimation methods (ML and GLS) affect SEM fit indices under both correctly-specified and misspecified models; and
4. to assess how sample size influences SEM fit indices and parameter estimates.

Method

SEM Fit Indices Studied

As with most studies in this area, the behaviors of χ^2 statistic (P-CHI) and the adjusted χ^2 statistic (P-ACHI) (i.e., χ^2 test corrected for elliptical distribution, a symmetrical distribution with uniform kurtosis; see Wang et al., 1996, and Browne, 1982 for more details) were examined in the present study. Although a variety of other SEM fit indices are available, some of

them are not readily comparable with each other. For example, Akaike's information criterion (AIC) has such a different metric from many other fit indices, and it is used in such a different fashion, that a meaningful comparison between AIC and GFI is difficult.

Based on the consideration of comparability, eight well-known SEM fit indices were chosen for investigation in the present study: goodness-of-fit index (GFI), adjusted goodness-of-fit index (AGFI), Bentler's comparative fit index (CFI), McDonald's centrality index (CENTRA), Bentler and Bonnett's non-normed fit index (N_NFI) and normed fit index (NFI), Bollen's normed fit index rho1 (RHO1), and Bollen's non-normed index delta2 (DELTA2). The GFI, AGFI, CFI are normed indices ranging from 0 to 1 in value, while non-normed indices can have values from 0 to slightly over 1. Of these eight fit indices, five of them belong to the category of <u>comparative model fit</u> indices (CFI, N_NFI, NFI, RHO1, and DELTA2) discussed before. Because parsimonious type of fit indices (James et al., 1982; Mulaik et al., 1989) are useful for assessing competing models, and they are not on the scale comparable with the eight indices above, they were not included in the study.

<u>Design of Monte Carlo Simulation</u>

Four factors were incorporated into the design of the study: data normality condition (three levels: normal, slightly nonnormal, and moderately nonnormal data), model specification (three levels: true, slightly misspecified, and moderately misspecified models), estimation methods (two levels: ML and GLS),

and sample size (four levels: 100, 200, 500, and 1000). The four factors were fully crossed with each other, creating 72 (3x3x2x4) different conditions. Within each condition, 200 replications were implemented to an acceptably small standard error of simulation. This balanced experimental design allows for a systematic assessment of the impact of the four factors on SEM fit indices and parameter estimates. The design required the generation of 14,400 random samples ($3 \times 3 \times 2 \times 4 \times 200$).

A widely-known model from substantive research (Wheaton, Muthén, Alwin, & Summers, 1977) with six observed and three latent variables was used in the simulation. This model has been discussed extensively in SEM literature (e.g., Bentler, 1992; Jöreskog & Sörbom, 1989). As suggested by Gerbing and Anderson (1993), simulating substantively meaningful models in Monte Carlo studies may increase the external validity of Monte Carlo research results. The true model with population parameters (presented in LISREL convention) and the two misspecified models are presented in Figure 1.

Insert Figure 1 about here

Although the population parameters presented in Figure 1 were arbitrarily specified, these parameters were specified to be close to the values in the original substantive research example so as to increase the external validity of the simulation results of the present study. Once the population parameters were fully specified, the population covariance matrix (Σ) was obtained

through the following formula (Jöreskog & Sörbom, 1989, p. 5), and this population covariance matrix was used to generate the random samples in the simulation:

$$\Sigma = \begin{bmatrix} \Lambda_y (I-B)^{-1} (\Gamma\phi\Gamma' + \psi) (I-B')^{-1} \Lambda_y' + \theta_\varepsilon & \Lambda_y (I-B)^{-1} \Gamma\phi\Lambda_x' \\ \Lambda_x\phi\Gamma' (I-B')^{-1} \Lambda_y' & \Lambda_x\phi\Lambda_x' + \theta_\delta \end{bmatrix}$$

Model Misspecification

Although a true model is relatively easy to specify in simulation research, model misspecification is difficult to handle for at least two reasons: (a) model misspecification can take such a variety of forms; and (b) the degree of model misspecification is not easily quantified. In other words, it is difficult to make *a priori* predictions about the severity of misspecification (Gerbing & Anderson, 1993). In the present study, model misspecification was achieved by fixing/constraining certain parameters in the model which should be free. The degree of model misspecification was empirically determined by fitting misspecified models to the population covariance matrix, and the resultant values of fit indices were used as indicators of severity of model misfit.

As the operational guideline, the "slightly misspecified" condition was defined as producing fit indices around .98 (for those approximately on the scale of 0 to 1) when the misspecified model was fit to the population covariance matrix, and a χ^2 test that would reach statistical significance for a sample size around 500. The "moderately misspecified" condition was defined as producing fit indices between .93 and .95 when the misspecified

model was fit to the <u>population</u> covariance matrix, and a χ^2 test that would reach statistical significance for a sample size around 100.

Obviously, the terms "slightly misspecified" and "moderately misspecified" are used here exclusively to indicate different degrees of misspecification, and by no means should these terms be generalized beyond this particular usage or beyond the present study. The two misspecified models are also presented in Figure 1.

<u>Data Nonnormality Conditions</u>

Similar to the issue of model misspecification, the degree of data nonnormality is not easily characterized in research. In other words, the criteria that can be used to deferentiate <u>slight</u>, <u>moderate</u>, and <u>severe</u> data nonnormality are not entirely clear. In the present study, the two data nonnormality conditions were specified *a priori* as follows: (a) for the “slightly nonnormal” condition, two thirds (2/3) of the observed variables have univariate skewness at about ±1.0, and univariate kurtosis at about ±1.0; (b) for the “moderately nonnormal” condition, two thirds (2/3) of the observed variables have univariate skewness at about ±1.5, and univariate kurtosis between +3 to +4. Again, such operational definitions should under no means be construed as representing rigid criteria; instead, the definitions should be treated simply as vehicles for operationally communicating the design protocol we employed.

Table 1 presents the population covariance matrix (correlations plus means and standard deviations) used for data

generation. Because the means of the variables do not affect SEM model fitting (unless a means model is tested), all the measured variables were centered with means being zeros so as to simplify the data generation process. The two data nonnormality conditions are also presented in Table 1.

Insert Table 1 about here

Data Generation

Data generation was accomplished using the data generator under the SAS System. To create each of the 14,400 sample data sets, the following steps were implemented:

1. six random normal variables with a desired sample size were generated, using the pseudorandom number generator under SAS;
2. the multivariate normality and nonnormality conditions were simulated using the matrix decomposition procedure (Kaiser & Dickman, 1962);
3. multivariate nonnormality conditions were simulated using:
   a. the power transformation method (Fleishman, 1978);
   b. the intermediate correlation procedure (Vale & Maurelli, 1983); and finally,
   c. the matrix decomposition procedure (Kaiser & Dickman, 1962);
4. the six correlated variables were linearly transformed to have desired means and standard deviations; and
5. The multivariate sample data were fitted to one of the models (true, slightly misspecified, and moderately misspecified)

under one of the two estimation procedures (ML and GLS), using PROC CALIS procedure under SAS. All desired fit indices and parameter estimates from each sample were obtained and saved for later analysis.

Simulation programming was implemented through a combination of the SAS Macro language, the SAS PROC IML matrix language, and the SAS PROC CALIS procedure for SEM model fitting under the SAS environment. All simulation was implemented on an IBM PC Pentium 100 MHZ computer with SAS Windows Version 6.11.

Results and Discussion

Convergence Failures and Improper Solutions

In simulation work involving SEM, it is normal to encounter two problems: the problem of nonconvergence, and that of improper solutions. The problem of nonconvergence occurs when SEM estimation fails to converge on a solution for a sample. The problem of improper solution occurs when some statistically impossible values, such as negative residual variances ("Heywood cases") are obtained from the estimation.

The problem of convergence failure in SEM depends to a great extent on the optimization procedure used and the number of iterations allowed for such optimization. Without information on the optimization procedure used and the number of iterations allowed, any discussion about convergence failure problems would be incomplete. In the present study, the Levenberg-Marquardt optimization technique was used, which is believed to work well for poor initial values. For discussion about this technique and additional references, readers are referred to SAS Institute

(1990, Chapter 14, pp. 245-366). Table 2 presents the percentage of non-convergent samples under different numbers of iterations, different sample size conditions, different estimation procedures (ML and GLS), under each of the three models (true, slightly misspecified, and moderately misspecified models), and under three data normality conditions.

Insert Table 2 about here

The results in Table 2 suggest five conclusions. First, as expected, convergence failures occurred mainly when the maximum number of iterations allowed was small. When the number of permitted iterations was increased to 40 and 50, convergence failure was rarely a problem. Second, also as expected, convergence failure was mainly a problem with small sample sizes. For example, under the sample size condition of 100 and the maximum number of iterations of 20, approximately 3.1% of the samples failed to converge. For the sample size of 200 with the same maximum number of iterations, only 0.53% of the samples failed to converge. When sample size reached 500, no convergence failures occurred.

Third, convergence failures appeared to occur more often under ML than under GLS estimation, with the ratio being approximately 2 to 1. Fourth, when the maximum numbers of iterations were relatively small, convergence failure appeared to occur substantially more often for moderately misspecified model than for the other two models. This makes intuitive sense in that

misspecified models may require larger number of iterations to reach optimal solutions. When the number of iterations allowed was increased, however, convergence failure became a negligible problem for all the three models. Fifth, data normality condition did not seem to influence estimation convergence in any systematic fashion, i.e., the occurrence of convergence failure did not depend on whether data were normal or nonnormal.

Table 3 presents the percentages of improper solutions under four factors: sample size, estimation methods, model specification, and data normality conditions. Again, it can be seen that improper solution is mainly a problem for smaller sample sizes. For example, for the sample size of 100, as many as 12.5% of the samples yielded improper solution of some kind. For the sample size of 200, the percentage dropped to 2.5%. When the sample size reached 500, this problem was practically eliminated. The two SEM estimation methods appeared to have a roughly equal percentage of improper solutions. Nonnormal data did not cause any more improper solutions than normal data.

Insert Table 3 about here

The findings that small sample size may often lead to convergence failure and/or improper solution in SEM were consistent with the findings of some previous studies (Anderson & Gerbing, 1984; Boomsma, 1985; Gerbing & Anderson, 1985), although the previous studies in this area mainly examined confirmatory factor analysis models rather than full structural equation

models. In addition to the findings related to sample size, the present study also extends exploration into new territories in examining several other factors potentially related to convergence failure and improper solutions in SEM, such as estimation methods, data nonnormality, and model misspecification. We are not aware that any previous studies have examined these issues.

Although both convergence failure and improper solution problems have long been identified in SEM simulation work, it is unclear how these two problems can best be handled in practice: to ignore them, to exclude them from subsequent analysis, or to replace them by generating new samples. In our study, we used 50 as the maximum number of iterations for each sample, which practically eliminated the problem of convergence failures, as shown in Table 2. For samples with improper solutions, we simply excluded these samples from subsequent analysis. As a result, the number of usable samples for analysis was reduced to 13,850 from the original 14,400, and the design of the experiment became slightly unbalanced.

<u>The Robustness of χ^2 and Adjusted χ^2 Tests</u>

In SEM applications, a major concern for the χ^2 test is its validity when data are nonnormal. Previous studies in this area indicated that the χ^2 test could be reasonably robust to nonnormal data conditions (e.g., Chou et al., 1991; Hu et al., 1992). The concern with nonnormal data also lead to the adjusted χ^2 test, which is the χ^2 test corrected for elliptical distribution, a symmetrical distribution with uniform kurtosis. Mathematically, the adjusted χ^2 statistic is obtained by dividing the χ^2 statistic

with the multivariate relative kurtosis coefficient (Browne, 1982, cited in SAS Institute, 1990, p. 305).

Table 4 presents the empirical rejection rates for the true model at the conventional α=.05 level. Under the normal data condition, the χ^2 and the adjusted χ^2 tests yielded almost identical rejection rates, and very close to the nominal probability level (α=.05). As the data became moderately nonnormal, the regular χ^2 test still yielded rejection rates very close to the nominal probability level even for the largest sample size of 1,000, while the adjusted χ^2 test yielded rejection rates considerably lower than the nominal α level. Furthermore, both ML and GLS estimation yielded very comparable rejection rates under all data normality and sample size conditions.

Insert Table 4 about here

The results in Table 4 indicate that, if we are concerned about the rejection or retention of the <u>true</u> model, the χ^2 test is quite robust to moderate data nonnormality (as defined in this study) even for sample sizes of 500 and 1000. The adjusted χ^2 test may be unnecessary for these data nonnormality conditions, because its correction seems to cause consistently lower empirical rejection rates than the nominal significance level. These results are generally consistent with findings in this area (e.g., Chou et al, 1991; Hu et al. 1992). But to what degree such robustness of the χ^2 test will hold under more severe nonnormality conditions is a question that needs to be addressed empirically.

Descriptive SEM Fit Indices

Because descriptive fit indices are designed to provide information about how well a model fits empirical data, and are not designed to provide information about sample size, data nonnormality, or the estimation techniques used for model fitting, it is almost self-evident that, ideally, a fit index (a) should be affected by the degree to which a model is incorrectly specified; (b) should not be unduly affected by data normality condition; (c) should not be unduly affected by the estimation method for model fitting; and (d) should not be unduly affected by sample size. In other words, the major factor contributing to the variation of an ideal fit index should be the model specification, and all the other three factors (data normality condition, estimation method, and sample size) should contribute minimally to variations in fit. Table 5 presents the results of partitioning the variance of the fit indices into different sources. Such variance partitioning allows systematic examination of the influences of the four factors discussed above.

Insert Table 5 about here

Under the initial balanced design of the study, variances contributed by different sources could have been partitioned orthogonally. In other words, variances due to different sources and their interactions would have been additive, which would have made interpreting the variance partitioning results very straightforward. However, due to the exclusion of the samples

with improper solutions, the design became slightly unbalanced. But this slight imbalance left the additive nature of the partitioned variances still reasonably intact.

Model specification. Although model specification indeed contributed most to the variation of all the fit indices examined in Table 5, the amount of variation accounted for by the model specification varied substantially among the fit indices, ranging from the high of 73% to the low of 35%. GFI and CENTRA had the highest proportion of variation accounted for by model specification (>70%), and RHO1 and N-NFI had the lowest amount of variation accounted for by this factor (35% and 45%). Viewed from the perspective that model specification should be the major contributor to the variation of an ideal fit index, it appears that GFI and CENTRA were the best two among the eight fit indices.

Data nonnormality. The factor of data normality condition turned out to be a nonevent, with no effect on any of the fit indices examined here. This is shown by the near zero proportions of variation that was accounted for by this factor for all the indices. Also, data normality as a factor was not involved in any meaningful interaction terms in the analysis either. These results indicate that all these fit indices were reasonably robust to the data nonnormality conditions as implemented in the present study.

We consider the degree of data nonnormality implemented in the study to have been somewhat mild, and it is not known from the present results whether this robustness to data nonnormality will hold under more severe nonnormality conditions. Quite a few fit

indices examined here are related to the χ^2 statistic in some fashion, it is possible that data nonnormality conditions not severe enough to cause misbehavior in the χ^2 statistic would not cause any misbehavior in these fit indices either. It will be interesting to see how these fit indices will behave under data nonnormality conditions severe enough to cause problems for the χ^2 statistic.

Estimation methods. The susceptibility of the eight indices to the influence of estimation methods varied considerably. The indices CFI, N_NFI, NFI, RHO1, and DELTA2 were strongly influenced by the estimation method used for model fitting (ML and GLS in this study), with 10% to 26% of variation accounted for by the estimation factor. It is interesting to note that all these five indices are comparative model fit indices. Based on the criterion that estimation method should not unduly influence an ideal fit index, it appears that the category of comparative model fit indices fared less well than the other three fit indices (GFI, AGFI, and CENTRA).

In Table 6, only one two-way interaction term (MS * EM: Model Specification * Estimation Method) is listed, because this is the only interaction term which accounted for noteworthy variations in some fit indices (CFI: 12%; N-NFI: 10%; NFI: 11%; RHO1: 8%; and DELTA2: 12%). All other two-, three-, and four-way interaction terms were not listed in the table because each of them accounted for a negligible portion of variation (< 1%) for any of the fit indices.

The strong interaction term between model specification and

estimation method for some fit indices indicates that the influence of estimation method on these fit indices is not uniform under the three fitted models (true, slightly misspecified, and moderately misspecified models). To fully understand these dynamics, a separate ANOVA was conducted to partition the variation of the fit indices under each of the three models, and the results are presented in Table 6.

Insert Table 6 about here

As discussed before, ideally, factors other than model specification should minimally contribute to the variation of a fit index. Under the same model, we would expect random variation to be the dominant source of variation for the indices, rather than any other factor or factors. The data presented in Table 6 show that under the true model, most fit indices performed well in this regard, and estimation method accounted for very small proportions of the variation for the indices, except for the CFI (2.91%) and NFI (14.25%) indices.

But as model misspecification became more severe, all those indices classified as comparative model fit indices were increasingly influenced by estimation method. For example, under the moderately misspecified model, estimation method was the dominant source of variation for these indices, accounting for up to 70% of the variation for some indices. On the other hand, the GFI and AGFI indices still remained immune to the influence of estimation method, and CENTRA was only slightly influenced by

estimation method.

Some previous studies (e.g., Wang et al., 1996) have concluded that these fit indices performed consistently and comparably under both ML and GLS estimation method. But the analysis presented here indicates that this and similar conclusions concerning estimation method may need to be revisited. It is shown here that, although the comparative model fit indices may indeed be comparable under either ML or GLS estimation method for the true model, such may not be the case for misspecified models, even when the degree of model misspecification is not severe, as was the case in the present study.

Sample size. In previous Table 5, sample size accounted for a considerable portion of variation of a few indices, including the GFI (7%), AGFI (14%), RHO1 (8%), and NFI (4%). This indicates that these indices are susceptible to the influence of the sample sizes used in SEM analysis. The practical implications of this influence will be further explored momentarily. Sample size had little influence on the CFI, CENTRA, N-NFI, and DELTA2 indices, which therefore performed well under the criterion that a fit index should not be unduly influenced by sample size.

To further understand the practical impact of estimation method and sample size on some of these fit indices, the descriptive statistics for these indices under two estimation methods and under different sample size conditions are presented in Table 7. For the sake of simplicity, we presented only basic descriptive information here (i.e., means and standard deviations).

Insert Table 7 about here

A close look at Table 7 reveals several phenomena. First, for all the three models, some fit indices (GFI, AGFI, RHO1, and NFI) exhibited a slight downward bias for smaller sample size conditions such as 100 and 200. Not surprisingly, these are indices for which sample size accounted for a considerable portion of their variation, as reported in Table 5 and as noted previously. In some cases, the magnitude of the downward bias may have practical implications for assessing model fit.

For example, for the true model (Model 1 in Table 7), the population parameters of AGFI was 1.00. For sample size of 100, the average AGFI was only .94. Similar downward bias was seen for RHO1, and to a lesser degree, for GFI and NFI. Practically, the existence of such downward bias indicates that when sample size is relatively small, researchers can hardly expects a value close to 1.00 for these indices, even if a perfect model has unknowingly been specified. The problem, of course, is that the applied researchers will not know whether the attenuated fit index is due to the bias caused by small sample size, or model misspecification, or both causes.

Second, under the true model, the population fit indices are identical under the two estimation methods (ML and GLS). For samples, these fit indices are either identical or very close to each other, except in a few cases where sample size is relatively small, such as for NFI and RHO1. But under misspecified models,

discrepancies between the two estimation methods (ML and GLS) occurred for some fit indices. The discrepancies became more conspicuous as the model misspecification became worse.

For the moderately misspecified model, which itself may not be considered as a bad model by most conventional standards, the discrepancy between ML and GLS fit indices became so large for some fit indices that they might lead to very different conclusions regarding model fit. For example, for N_NFL (.91 vs. .74), NFI (.95 vs. .85), RHO1 (.90 vs. .72), if the interpretation was based on ML fit indices, the model would most probably be said to have reasonable, though not great, fit with data. But if the interpretation was based on GLS fit indices values, it is very likely that the model would be considered to have very poor fit. This phenomenon has not been widely discussed in the literature, although it has been previously noted (Fan et al., 1996).

Third, as discussed in previous sections, those fit indices which exhibit discrepancy between ML and GLS methods were all <u>comparative model fit</u> indices (CFI, N_NFI, NFI, RHO1, and DELTA2). The other fit indices (GFI, AGFI, and CENTRA), which do not rely on the comparison between a fitted model and a more restricted null model, showed remarkable consistency between ML and GLS methods under all three model specification conditions. Although the reasons for this phenomenon are not entirely clear to us, this descriptive information confirms the observation from the variation partitioning analysis presented in Tables 5 and 6, where estimation method turned out to account for a considerable portion of variation for only selected indices.

These results are also very consistent with those from a similar study (Fan et al., 1996) which involved a different SEM model. These findings lead us to believe that comparative model fit indices in general are more susceptible to the influence of estimation methods, and as a result, the interpretation of such indices may be more uncertain under the two normal theory estimation methods.

It is probably safe to say that, in research practice, there does not exist any true SEM model, because a true model is more a mathematical abstraction than a reality. As a result, the question is not whether the fitted model is a true model, but rather, how well the model approximates the data (Bentler & Dudgeon, 1996; Cudeck & Henly, 1991). In this sense, it is the model with some degree of misspecification that researchers have to make decisions about in their applied research.

The two misspecified models examined in the present study probably represent the *least* degrees of misspecification that applied researchers may encounter in practice. The slightly misspecified model examined here would probably be judged as having very good model fit by any current conventional criteria. Even the moderately misspecified model would be regarded as having reasonable model fit by most conventional criteria. In this context, for the misspecified models in the present study, the discrepancies exhibited by some fit indices under the two estimation methods must be considered very disturbing.

For example, under the moderately misspecified model, the five comparative model fit indices (CFI, N-NFI, NFI, RHO1, DELTA2)

all showed differences of about 0.1 or larger between ML and GLS estimation methods, with RHO1 having the largest discrepancy (0.18; ML: 0.90; GLS: 0.72). When used for assessing model fit, a discrepancy of 0.10 near the upper ceiling of the fit index value may lead to quite different conclusions about model fit.

Using the criterion that a fit index should be influenced by model specification, but not unduly influenced by confounding factors such as data nonnormality, estimation method, or sample size, it appears that the CENTRA index was the top performer among the indices investigated here, and followed by GFI. Other indices were strongly susceptible to the influence of one or more of the confounding factors we investigated. This finding that CENTRA has outstanding performance (followed by GFI) is consistent with findings from a previous study involving a different SEM model (Fan et al., 1996).

Data Nonnormality and Parameter Estimates

In addition to the SEM fit indices, the potential effect of data nonnormality on the quality of the SEM parameter estimates has also been an important concern (e.g., Wang et al., 1996). Afterall, even when we can correctly identify degree of model fit, we then want to examine the parameter estimates to evaluate the substantive meaning of the model.

The major question asked in the context of this second issue is whether and to what degree the quality of parameter estimates in SEM will be adversely affected when data normality assumption in SEM is violated. Table 8 presents the mean estimates for the 17 parameters in the model. Due to space considerations, we were

not able to present all the data. In Table 8, what are presented are estimates based on maximum likelihood estimation, for the true and moderately misspecified models, for normal data and moderately nonnormal data conditions, and for sample size of 100, 500, and 1000.

The mean estimates presented in Table 8 gave no indication of any systematic adverse effects that data nonnormality might have on the quality of these parameter estimates when compared with those estimates under the normal data condition for both the true and moderately misspecified models. In other words, the mean estimates under normal data conditions are not necessarily more accurate than those under nonnormal data conditions; any discrepancies appear to be random rather than systematic.

This indicates that for the nonnormal data conditions implemented in this study, the adverse effect of data nonnormality, if any, may be so minor that it may not cause much concern for the quality of mean parameter estimates. However, as discussed previously, the degree of data nonnormality implemented in the study was not especially severe. So it is not known if the robustness of parameter estimates as seen here will hold under more severe nonnormal data conditions.

To provide a more systematic assessment of any potential effect of data nonnormality on parameter estimates, variance partitioning was also applied to the 17 parameter estimates to check what factors contributed to the variation of each estimate in repeated sampling. Because population parameters differed across the three models (true, slightly misspecified, and

moderately misspecified models), the variance partitioning was carried out separately under each of the three models for all the 17 parameters.

If the data normality condition had any systematic effect on the parameter estimates, this would be reflected in this analysis as a strong data normality factor accounting for a substantial portion of variation in the parameter estimates, or as a strong interaction term involving data normality condition as one factor in the interaction. This variance partitioning analysis required carrying out 51 analyses of variance (17 parameters under each of the three models). The results across parameters and across models invariably showed that data normality condition was a factor accounting for much less than one percent of the variation in the each of the parameter estimates. Furthermore, no interaction term involving data normality condition was observed to account for any noteworthy portion of variation of the parameter estimates. Thus, both the mean parameter estimates in Table 8 and the variance partitioning analyses for the parameter estimates indicated that data nonnormality condition has no discernible effect on the quality of SEM parameter estimates. Again, it remains an empirical question whether this view will hold under more severe data nonnormality conditions.

Limitations and Future Directions

As with most empirical studies, the present study had its own share of limitations. The most obvious limitation is that there was only one model simulated, thus the findings may reflect some idiosyncracies associated with the model, and the study does not

provide any mechanisms for verification of these findings except by comparison with previous studies which also considered model misspecifucation (Fan et al., 1996). This limitation can be offset if a series of similar studies can be conducted which involve additional different SEM models.

Another potential limitation of the study is that we were not always able to provide theoretical rationales for some phenomena observed in the study. Although it is desirable to have theoretical explanations for empirically observed phenomena, this is not always possible in all aspects of SEM simulation or analysis.

Regarding future research, as we have emphasized throughout this paper, we believe that it is important that SEM simulation must involve not only correctly specified models, but also models with some degree of misspecification. Otherwise, SEM simulations will have little ecological validity as regards applied research. Indeed, model fit indices may behave quite differently under models with even only minor degrees of misspecification, and these dynamics may be more important for us to understand than the behaviors of the same indices under a mathematically perfect specification representing an unattainable ideal.

For this reason, future research in this area should consider different aspects of SEM analysis under realistically misspecified SEM models, instead of focusing solely on the true SEM model. Of course, future research will benefit from incorporating several different SEM models with different degrees of model complexity in one study so that the chances of fluke results can be reduced, and

more meaningful findings can be realized.

Summary and Conclusions

An experimental design was used in the present empirical study to investigate the effects of four factors on SEM fit indices and parameter estimates. Under this experimental design, a total of 14,400 samples were generated and fitted to three SEM models with different degrees of model misspecification. The effects of data nonnormality, estimation method, and sample size on SEM fit indices and the effect of data nonnormality on parameter estimates were systematically assessed. The major findings were:

1. In SEM model fitting, the problems of convergence failure and improper solutions are associated with smaller sample sizes. If the number of iteration allowed is not too restricting, convergence failure appeared to be a negligible problem. Improper solutions, on the other hand, seems to be a more serious issue, especially when sample size is small. Other factors, such as data nonnormality and model specification, do not seem to be related to these two problems.
2. When the degree of data nonnormality is mild or even slightly moderate, the χ^2 test may be quite robust in the sense that the empirical rejection rate of the true model is very close to the nominal alpha level, even when the sample size is moderately large.
3. Data nonnormality does not systematically affect the eight descriptive SEM fit indices examined in any discernible fashion. Although under the true model, the eight fit

indices were quite consistent under the two normal theory estimation methods, under the two misspecified models the estimation method exhibited considerable influence on some fit indices. Specifically, all the fit indices belonging to the category of comparative model fit indices tended to be affected to noteworthy degress by the estimation method used for model fitting. As the degree of model misspecification increased, these discrepancies become sufficiently large to lead to quite different interpretations regarding SEM model fit.

4. Sample size had considerable influence on a few indices, and these were the indices with an obvious tendency of downward bias under smaller sample size conditions. This downward bias could have some very real practical implications in applied research.
5. Data nonnormality conditions as implemented in the study had very little adverse effect on the quality of SEM parameter estimates. Although the data nonnormality conditions implemented in the study were not extremely severe, these results gave some indication that the SEM parameter estimation process is robust to mild to moderate data nonnormality.

Overall, the effect of data nonnormality appears to be rather weak, or even nonexistent, for both SEM model fit assessment (the χ^2 statistic and descriptive fit indices) and SEM parameter estimation, and SEM analysis appears to be quite robust against mild or even moderate deviations from normality. Given that the

data nonnormality conditions implemented in the study were somewhat mild, we were not expecting strong adverse effects of data nonnormality. But the almost complete absence of any obvious adverse effect of data nonnormality was still somewhat surprising to us. This finding may somewhat alleviate obsessive concerns about data nonnormality in SEM application; of course, this does not imply that such issues can be ignored.

Among the descriptive SEM fit indices, the centrality index performed best, followed by the goodness-of-fit index (GFI). This result is consistent with the findings from a previous study (Fan et al., 1996). Although the finding must still be regarded as somewhat tentative, the remarkable consistency across the two studies involving different SEM models has appreciably increased our confidence that the finding is replicable and noteworthy.

References

Anderson, J. C., & Gerbing, D. W. (1984). The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. Psychometrika, 49, 155-173.

Bearden, W. O., Sharma, S., & Teel, J. R. (1982). Sample size effects on chi-square and other statistics used in evaluating causal models. Journal of Marketing Research, 19, 425-530.

Bentler, P. M. (1983). Some contributions to efficient statistics for structural models: Specification and estimation of moment structures. Psychometrika, 48, 493-517.

Bentler, P. M. (1990). Comparative fit indexes in structural models. Psychological Bulletin, 107, 238-246.

Bentler, P. M. (1992). EQS structural equations program manual. Los Angeles, CA: BMDP Statistical Software.

Bentler, P. M. (1994a). Forward. In B. M. Byrne, Structural Equation Modeling with EQS and EQS/Windows. Thousand Oaks, CA: SAGE Publications.

Bentler, P. M. (1994b). On the quality of test statistics in covariance structure analysis: Caveat emptor. In C. R. Reynolds (Ed.), Cognitive assessment: A multidisciplinary perspective (pp. 237-260). New York: Plenum.

Bentler, P. M., & Dudgeon, P. (1996). Covariance structure analysis: Statistical practice, theory, and directions. Annual Review of Psychology, 47, 563-592.

Bagozzi, R. P., Fornell, C., & Larcker, D. F. (1981). Canonical correlation analysis as a special case of structural

relations model. Multivariate Behavioral Research, 16, 437-454.

Baldwin, B. (1989). A primer in the use and interpretation of structural equation models. Measurement and Evaluation in Counseling and Development, 22, 100-112.

Bollen, K. A. (1986). Sample size and Bentler and Bonnett's nonnormed fit index. Psychometrika, 51, 375-377

Bollen, K. A. (1989). A new incremental fit index for general structural equation models. Sociological Methods & Research, 17, 303-316.

Bollen, K. A. & Long, J. S. (1993). Introduction. In K. A. Bollen and J. S. Long (Eds.), Testing structural equation models (pp. 1-9). Newbury Park, California: SAGE Publications.

Bollen, K. A., & Stine, R. (1992). Bootstrapping goodness-of-fit measures in structural equation models. Sociological Methods and Research, 21, 205-229.

Boomsma, A. (1982). The robustness of LISREL against small sample sizes in factor analysis models. In K. G. Jöreskog & H. Wold (Eds.), Systems under indirect observation: Causality, structure, prediction (Part I) (pp. 149-175). Amsterdam: North-Holland.

Boomsma, A. (1985). Nonconvergence, improper solutions, and starting values in LISREL maximum likelihood estimation. Psychometrika, 50, 229-242

Boomsma, A. (1987). The robustness of maximum likelihood estimation in structural equation models. In P. Cuttance & R. Ecob (Eds.), Structural modeling by example: Applications in educational, sociological, and behavioral research (pp. 160-188).

New York: Cambridge University Press.

Browne, M. W. (1982). Covariance structures. In D. M. Hawkins (Ed.), Topics in applied multivariate analysis (pp. 72-141). Cambridge, England: Cambridge University Press.

Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. British Journal of Mathematical and Statistical Psychology, 37, 62-83.

Byrne, B. M. (1994). Structural equation modeling with EQS and EQS/Windows: Basic concepts, applications, and programming. Newbury, California: SAGE Publications, Inc.

Camstra, A., & Boomsma, A. (1992). Cross-validation in regression and covariance structure analysis: An overview. Sociological Methods & Research, 21, 89-115.

Chou, C. P., Bentler, P. M., & Satorra, A. (1991). Scaled test statistics and robust standard errors for nonnormal data in covariance structure analysis: A Monte Carlo study. British Journal of Mathematical and Statistical Psychology, 44, 347-357.

Cudeck, R., & Henly, S. J. (1991). Model selection in covariance structure analysis and the "problem" of sample size: A clarification. Psychological Bulletin, 109, 512-519.

Fan, X. (1996). Structural equation modeling and canonical correlation analysis: What do they have in common? Structural Equation Modeling: A Multidisciplinary Journal, 4, 64-78.

Fan, X., Wang, L. & Thompson, B. (1996, April). The effects of sample size, estimation methods, and model misspecification on SEM fit indices. Paper presented at the annual meeting of American Educational Research Association, New York. (ERIC

Document Reproduction Service No. ED forthcoming)

Fleishman, A. I. (1978). A method for simulating non-normal distributions. Psychometrika, 43, 521-531.

Gerbing, D. W., & Anderson, J. C. (1985). The effects of sampling error and model characteristics on parameter estimation for maximum likelihood confirmatory factor analysis. Multivariate Behavioral Research, 20, 255-271.

Gerbing, D. W., & Anderson, J. C. (1993). Monte Carlo evaluations of goodness-of-fit indices for structural equation models. In K. A. Bollen & J. S. Long, (Eds.), Testing structural equation models (pp. 40-65). Newbury Park, CA: SAGE Publications, Inc

Hu, L., Bentler, P. M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? Psychological Bulletin, 112, 351-362.

Ichikawa, M., & Konishi, S. (1995). Application of the bootstrap methods in factor analysis. Psychometrika, 60, 77-93.

James, L. R., Mulaik, S. A., & Brett, J. M. (1982). Causal analysis: Models, assumptions, and data. Beverly Hills, CA: Sage.

Jöreskog, K. G. & Sörbom, D. (1989). LISREL 7: A Guide to the Program and applications, 2nd Edition. Chicago, IL: SPSS Inc.

Kaiser, H. F., & Dickman, K. (1962). Sample and population score matrices and sample correlation matrices from an arbitrary population correlation matrix. Psychometrika, 27, 179-182.

Kano, Y., Berkane, M., & Bentler, P. M. (1990). Covariance structure analysis with heterogeneous kurtosis parameters.

Biometrika, 77, 575-585.

La Du, T. J., & Tanaka, J. S. (1989). The influence of sample size, estimation method, and model specification on goodness-of-fit assessment in structural equation models. Journal of Applied Psychology, 74, 625-635.

Loehlin, J. C. (1992). Latent variable models: An introduction to factor, path, and structural analysis. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.

Maiti, S. S., & Mukherjee, B. N. (1991). Two new goodness-of-fit indices for covariance matrices with linear structure. British Journal of Mathematical and Statistical Psychology, 44, 153-180.

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. Psychological Bulletin, 103, 391-410.

MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. Psychological Bulletin, 111, 490-504.

McDonald, R. P. (1989). An index of goodness-of-fit based on noncentrality. Journal of Classification, 6, 97-103.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. Psychological Bulletin, 105, 156-166.

Mooijaart, A. (1985). Factor analysis for non-normal variables. Psychometrika, 50, 323-342.

Mulaik, S. A., James, L. R., Van Alstine, J., Bennett, N., Lind, S., & Stillwell, C. D. (1989). An evaluation of goodness of

fit indices for structural equation models. Psychological Bulletin, 105, 430-445.

Pedhazur, E. J. & Schmelkin, L. P. (1991). Measurement, design, and analysis: An integrated approach. Hillsdale, NJ: Lawrence Erlbaum Associates.

SAS Institute, Inc. (1990). SAS/STAT user's guide, version 6 (vol. 1, 4th ed.). Cary, NC: SAS Institute Inc.

Sobel, M. E., & Bohrnstedt, G. W. (1985). Use of null models in evaluating the fit of covariance structure models. In N. B. Tuma (Ed.), Sociological methodology 1985 (pp. 152-178). San Francisco, Jossey-Bass.

Tanaka, J. S. (1987). "How big is big?": Sample size and goodness-of-fit in structural equation models with latent variables. Child Development, 58, 134-146.

Tanaka, J. S. (1993). Multifaceted conceptions of fit in structural equation models. In K. A. Bollen & J. S. Long, (Eds.), Testing structural equation models (pp. 10-39). Newbury Park, CA: SAGE Publications, Inc.

Tanaka, J. S., Huba, G. J. (1989). A fit index for covariance structure models under arbitrary GLS estimation. British Journal of Mathematical and Statistical Psychology, 42, 233-239.

Thompson, B. (1996). AERA editorial policies regarding statistical significance testing: Three suggested reforms. Educational Researcher, 25(2), 26-30.

Thompson, B., & Daniel, L.G. (1996). Factor analytic evidence for the construct validity of scores: An historical overview and

some guidelines. Educational and Psychological Measurement, 56, 213-224.

Vale, C.D. & Maurelli, V.A. (1983). Simulating multivariate nonnormal distributions. Psychometrika, 48, 465-471.

Wang, L., Fan, X., & Willson, V. (1996). Effects of nonnormal data on parameter estimates and fit indices for a model with latent and manifest variables: An empirical study. Structural Equation Modeling: A Multidisciplinary Journal, 3, 228-247.

Wheaton, D. E., Muthén, B., Alwin, D. F., & Summers, G. F. (1977). Assessing reliability and stability in panel models. In D. R. Heise (Ed.), Sociological methodology 1977 (pp. 84-136. San Francisco: Jossey Bass.

Table 1 Population Covariance Matrix and Data Nonnormality Conditions

							Nonnormality Conditions			
							Slight		Moderate	
							Skew[a]	Kurt	Skew	Kurt
σ^2	4.25	4.25	12.27	9.28	12.91	9.81				
μ	0	0	0	0	0	0				
X1	1.000000						-1.00	1.00	-1.50	3.50
X2	.536875	1.000000					1.00	1.00	1.50	3.50
Y1	-.379164	-.290805	1.000000				1.00	1.00	1.50	3.50
Y2	-.414038	-.317552	.669246	1.000000			.25	-.25	.25	-.25
Y3	-.375884	-.288290	.569164	.482667	1.000000		-1.00	1.00	-1.50	3.50
Y4	-.388047	-.297619	.456315	.529719	.672364	1.000000	-.25	.25	-.25	.25

a Skew: skewness; Kurt: kurtosis

Table 2: Percentage of Non-converging Samples under Four Factors

	Maximum Numbers of Iteration Allowed				
	20	25	30	40	50
Sample Size					
100	3.06	1.56	0.83	0.33	0.19
200	0.53	0.17	0.06	0.00	0.00
500	0.00	0.00	0.00	0.00	0.00
1000	0.00	0.00	0.00	0.00	0.00
Estimation Methods[a]					
ML	1.15	0.57	0.32	0.10	0.07
GLS	0.64	0.29	0.13	0.07	0.03
Models[b]					
True	0.38	0.21	0.15	0.13	0.10
Slight Mis.	0.29	0.13	0.08	0.04	0.00
Moderate Mis.	2.02	0.96	0.44	0.08	0.04
Data Normality					
Normal	0.71	0.40	0.23	0.13	0.08
Slight Nonnormal	1.10	0.42	0.17	0.02	0.02
Moderate Nonnormal	0.88	0.48	0.27	0.10	0.04

a ML: maximum likelihood; GLS: generalized least squares

b True, slightly misspecified, and moderately misspecified model respectively.

Table 3: Percentage Samples with Improper Solutions under Four Factors (Maximum Number of Iterations = 50)

Sample Size	100	200	500	1000
	12.53	2.53	0.14	0.00
Estimation Methods[a]	ML	GLS		
	3.74	3.86		
Models[b]	True	Slight	Moderate	
	4.63	5.94	0.83	
Data Nonnormality[c]	Normal	Slight	Moderate	
	3.58	3.83	3.98	

a ML: Maximum likelihood; GLS: generalized least squares

b True, slightly misspecified, and moderately misspecified models respectively.

c Normal, slightly non-normal, and moderately non-normal data conditions

Table 4: Empirical Rejection Rate (%) of the True Model for α=.05

		Sample Size							
		ML				GLS			
Data Condition	Tests	100	200	500	1000	100	200	500	1000
Normal	χ^2 Test	5.6	5.6	6.5	4.0	4.2	4.7	5.0	3.0
	Adj. χ^2 Test	6.2	5.6	7.0	4.0	4.2	4.7	5.0	3.5
Slightly Nonnormal	χ^2 Test	3.6	1.5	5.0	5.0	3.0	5.1	4.0	5.0
	Adj. χ^2 Test	3.0	1.0	4.5	3.5	3.6	4.1	3.5	4.0
Moderately Nonnormal	χ^2 Test	4.1	6.2	6.5	5.0	3.7	3.1	7.5	5.5
	Adj. χ^2 Test	1.8	1.0	2.0	1.0	0.6	1.6	1.0	1.0

Table 5 Eta-Squares of Different Sources of Variation for the Fit Indices

Source	Fit Indices							
	GFI	AGFI	CFI	CENTRA	N-NFI	NFI	RHO1	DELTA2
Model Specification	72.73 [a]	52.62	53.80	71.30	44.53	52.22	35.26	55.67
Data Normality	.01	.01	.00	.01	.01	.00	.01	.01
Estimation Methods	.09	.12	12.77	.57	10.78	18.56	26.15	10.20
Sample Size	6.99	14.44	.01	.13	.16	3.99	7.98	.25
MS * EM	.09	.09	11.71	.64	9.96	10.91	7.13	11.68
. [b]								
.								
Random Variation	19.54	32.42	21.37	26.99	34.29	13.38	21.52	21.73

a Proportion of variance contributed by a source, obtained through:

η^2 = [(sum of squares due to a source)/(total sum of squares)] * 100

b The other five two-way interaction terms, four three-way interaction terms, and one four-way interaction terms all have η^2 less than one percent of the total sum of squares for all the fit indices. To avoid cluttering the table, these terms are omitted from the table.

Table 6 Percentage of Variation Accounted for by Estimation Method under Each Model

	Fit Indices							
Model	**GFI**	**AGFI**	**CFI**	**CENTRA**	**N-NFI**	**NFI**	**RHO1**	**DELTA2**
True	.00	.00	2.91	.00	.20	14.25	.16	.00
Slightly Misspecified	.18	.18	20.97	.63	13.92	36.02	36.01	14.46
Moderately Misscpecified	.96	.97	59.57	5.72	58.87	70.27	70.27	58.34

Table 7 Means and Standard Deviations of Ten Fit Indices under Two Estimation Methods

Indices[a]	GFI	AGFI	CFI	CENTRA	N_NFI	NFI	RHO1	DELTA2
Model 1[b]	1.00[c]	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Parameters	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Sample Sizes								
100	98 (01)[d]	94 (04)	100 (01)	100 (01)	100 (05)	98 (01)	94 (04)	100 (01)
	98 (01)	93 (05)	98 (03)	100 (01)	101 (15)	95 (03)	82 (11)	100 (03)
200	99 (00)	97 (02)	100 (00)	100 (01)	100 (02)	99 (01)	97 (02)	100 (01)
	99 (00)	97 (02)	99 (01)	100 (01)	100 (07)	97 (02)	90 (06)	100 (02)
500	100 (00)	99 (01)	100 (00)	100 (00)	100 (01)	100 (00)	99 (01)	100 (00)
	100 (00)	99 (01)	100 (01)	100 (00)	100 (02)	99 (01)	96 (03)	100 (01)
1000	100 (00)	99 (00)	100 (00)	100 (00)	100 (00)	100 (00)	99 (00)	100 (00)
	100 (00)	99 (00)	100 (00)	100 (00)	100 (01)	99 (00)	98 (01)	100 (00)
Model 2	.99	.96	.99	.99	.97	.99	.97	.99
Parameters	.99	.96	.97	.99	.92	.97	.90	.97
Sample Sizes								
100	97 (01)	90 (06)	98 (02)	98 (02)	97 (06)	97 (02)	91 (05)	99 (02)
	97 (01)	90 (05)	96 (04)	99 (02)	92 (17)	91 (04)	74 (13)	97 (05)
200	98 (01)	93 (04)	99 (01)	98 (01)	97 (04)	98 (01)	94 (03)	99 (01)
	98 (01)	94 (03)	97 (03)	99 (01)	91 (10)	94 (03)	82 (08)	97 (03)
500	99 (00)	95 (02)	99 (01)	98 (01)	97 (02)	98 (01)	95 (02)	99 (01)
	99 (00)	95 (02)	97 (02)	99 (01)	90 (06)	95 (02)	87 (06)	97 (02)
1000	99 (00)	96 (01)	99 (00)	98 (01)	97 (01)	98 (00)	96 (01)	99 (00)
	99 (00)	96 (01)	97 (01)	99 (00)	90 (04)	96 (01)	88 (04)	97 (01)

(To be continued)

Table 7 Means and Standard Deviations of Ten Fit Indices under Two Estimation Methods (Continued)

Indices	GFI	AGFI	CFI	CENTRA	N_NFI	NFI	RHO1	DELTA2
Model 3	.96	.90	.95	.94	.91	.95	.90	.95
Parameters	.96	.90	.86	.95	.74	.85	.72	.86
Sample Sizes								
100	94 (02)	84 (05)	95 (03)	94 (04)	90 (06)	92 (03)	85 (06)	95 (03)
	94 (02)	85 (05)	86 (08)	96 (03)	74 (15)	79 (06)	60 (11)	88 (07)
200	95 (01)	87 (04)	95 (02)	94 (02)	90 (04)	93 (02	87 (04)	95 (02)
	95 (01)	87 (04)	85 (05)	95 (02)	73 (09)	82 (05)	66 (08)	86 (05)
500	96 (01)	89 (03)	95 (01)	94 (02)	90 (03)	94 (01)	89 (03)	95 (01)
	96 (01)	89 (02)	85 (03)	95 (01)	72 (06)	83 (03)	69 (06)	85 (03)
1000	96 (01)	89 (02)	95 (01)	94 (01)	90 (02)	94 (01)	90 (02)	95 (01)
	96 (01)	90 (01)	85 (02)	95 (01)	72 (04)	84 (02)	71 (04)	85 (02)

a GFI: Goodness-of-fit index; AGFI: Adjusted goodness-of-fit index; CFI: comparative fit index; CENTRA: MacDonald's centrality index; N_NFI: Bentler-Bonnett's non-normed fit index; NFI: Bentler-Bonnett's normed fit index; RHO1: Bollen's Rho1; DELTA2: Bollen's delta2 (incremental fit) index; PGFI: James et al.'s parsimonious goodness-of-fit index; PARMS: Mulaik et al.'s parsimonious fit index.

b Model 1: True model; Model 2: slightly misspecified model; Model 3: moderately misspecified model.

c Population values of the fit indices; those in the upper rwo are based on maximum likelihood estimation, while those in the second row are based on generalized least squares estimation. These population values were obtained by fitting each model to the population covariance matrix with *N* set at 10,000.

d mean (standard deviation) of a fit index; the values in first row were based on maximum likelihood estimation, and those in the second row were based on generalized least squares estimation. Second place decimal point is omitted.

Table 8 Mean Estimates for Parameters (Under Maximum Likelihood Estimation)

Parameter Names	λ_1	λ_2	λ_3	γ_1	γ_2	β	ϕ	δ_1	δ_2	ϵ_1	ϵ_2	ϵ_3	ϵ_4	ζ_1	ζ_2	σ_{13}	σ_{24}
True Model Parameters	.50	.95	.90	-.60	-.25	.60	7.00	3.00	2.50	4.75	2.50	4.50	3.00	5.00	4.00	1.60	.30
Normal Data																	
N 100	.52	.97	.92	-.63	-.27	.60	7.14	3.09	2.45	4.65	2.47	4.51	2.82	4.74	3.78	1.65	.29
500	.50	.95	.90	-.60	-.26	.60	7.06	2.94	2.50	4.74	2.45	4.49	2.97	5.01	4.00	1.59	.29
1000	.50	.96	.90	-.60	-.25	.60	6.99	3.02	2.51	4.77	2.49	4.44	3.04	4.95	3.99	1.59	.31
Moderately Non-Normal Data																	
100	.54	.95	.89	-.62	-.29	.61	6.87	2.91	2.41	4.18	2.69	4.13	3.19	4.87	4.03	1.40	.39
500	.51	.97	.90	-.60	-.25	.61	6.96	2.99	2.46	4.74	2.44	4.47	3.01	4.97	4.00	1.60	.29
1000	.50	.96	.90	-.59	-.25	.61	7.07	2.96	2.50	4.76	2.44	4.46	2.93	4.94	3.99	1.62	.25
Mod. Misspe.Model[a] Parameters	.76	.76	.80	-.85	.	.79	4.50	4.82	1.92	3.70	3.54	3.49	3.73	5.83	3.77	.	.
Normal Data																	
100	.77	.77	.82	-.87	.	.79	4.70	4.68	1.86	3.71	3.42	3.48	3.67	5.65	3.70	.	.
500	.76	.76	.81	-.85	.	.79	4.49	4.82	1.92	3.70	3.54	3.50	3.65	5.83	3.75	.	.
1000	.76	.76	.80	-.84	.	.79	4.50	4.82	1.92	3.67	3.53	3.49	3.72	5.77	3.76	.	.
Moderately Non-Normal Data																	
100	.77	.77	.80	-.85	.	.82	4.52	4.84	1.87	3.65	3.52	3.25	3.85	5.78	3.80	.	.
500	.76	.76	.80	-.85	.	.79	4.52	4.84	1.90	3.72	3.54	3.40	3.73	5.77	3.81	.	.
1000	.77	.77	.80	-.85	.	.80	4.44	4.83	1.90	3.71	3.52	3.48	3.73	5.79	3.78	.	.

a Moderately Misspecified Model

$$\Lambda_x = \begin{bmatrix} 1.00 \\ 0.50 \end{bmatrix} \qquad \Lambda_y = \begin{bmatrix} 1.00 & 0.00 \\ 0.95 & 0.00 \\ 0.00 & 1.00 \\ 0.00 & 0.90 \end{bmatrix}$$

$$\Gamma = \begin{bmatrix} -0.60 \\ -0.25 \end{bmatrix} \qquad \Phi = [\ 7.00\ ]$$

$$B = \begin{bmatrix} 0.00 & 0.00 \\ 0.60 & 0.00 \end{bmatrix} \qquad \Psi = \begin{bmatrix} 5.00 & 0.00 \\ 0.00 & 4.00 \end{bmatrix}$$

$$\Theta_\delta = \begin{bmatrix} 3.00 & 0.00 \\ 0.00 & 2.50 \end{bmatrix} \qquad \Theta_\epsilon = \begin{bmatrix} 4.75 & & & \\ 0.00 & 2.50 & & \\ 1.60 & 0.00 & 4.50 & \\ 0.00 & 0.30 & 0.00 & 3.00 \end{bmatrix}$$

Figure 1 Models and True Model Parameters

TM026500

- AERA

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement (OERI)
Educational Resources Information Center (ERIC)

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: EFFECTS OF DATA NONNORMALITY AND OTHER FACTORS ON FIT INDICES AND PARAMETER ESTIMATES FOR TRUE AND MISSPECIFIED SEM MODELS

Author(s): XITAO FAN, LIN WANG and BRUCE THOMPSON

Corporate Source:

Publication Date: 3/25/97

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic/optical media, and sold through the ERIC Document Reproduction Service (EDRS) or other ERIC vendors. Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce the identified document, please CHECK ONE of the following options and sign the release below.

[X] Sample sticker to be affixed to document

Check here
Permitting microfiche (4"x 6" film), paper copy, electronic, and optical media reproduction

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
BRUCE THOMPSON
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Level 1

Sample sticker to be affixed to document [ ]

or here
Permitting reproduction in other than paper copy.

"PERMISSION TO REPRODUCE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY
Sample
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Level 2

Sign Here, Please

Documents will be processed as indicated provided reproduction quality permits. If permission to reproduce is granted, but neither box is checked, documents will be processed at Level 1.

"I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce this document as indicated above. Reproduction from the ERIC microfiche or electronic/optical media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries."

Signature:	Position: PROFESSOR
Printed Name: BRUCE THOMPSON	Organization: TEXAS A&M UNIVERSITY
Address: TAMU DEPT EDUC PSYC COLLEGE STATION, TX 77843-4225	Telephone Number: (409) 845-1831
	Date: 1/29/97

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC , *or*, if you wish ERIC to cite the availability of this document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents which cannot be made available through EDRS).

Publisher/Distributor:	
Address:	
Price Per Copy:	Quantity Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name and address of current copyright/reproduction rights holder:

Name:

Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

If you are making an unsolicited contribution to ERIC, you may return this form (and the document being contributed) to:

ERIC Facility
1301 Piccard Drive, Suite 300
Rockville, Maryland 20850-4305
Telephone: (301) 258-5500